数理解析研究所講究録
第2040巻 2017年 35-42

整列問題の平均比較回数

照山 順一
国立情報学研究所,
JST, ERATO, 河原林巨大グラフプロジェクト

岩間 一雄
京都大学数理解析研究所

Junichi Teruyama
National Institute for Informatics,
JST, ERATO, Kawarabayashi Large Graph Project

Kazuo Iwama
RIMS, Kyoto University

概要

比較ソートにおける効率は主に比較回数によって計られる．本研究では，できるだけ平均の比較回数が少ない比較ソートアルゴリズムの設計を目的とする．Edelkamp と Weißによって Ford と Johnson のアルゴリズムの平均比較回数が $n\lg n-1.3999n+O(\lg n)$ であることが示されている．また，情報理論的下限から平均比較回数の下限は $\lceil \lg n! \rceil \approx n\lg n-1.4426n+O(\lg n)$ であることが知られている．本研究では，2つの要素を挿入する操作を与えることにより，平均比較回数が高々$n\lg n-1.4034+O(\lg n)$ となるアルゴリズムを設計した．また，我々のアルゴリズムと Ford と Johnson のアルゴリズムを組み合わせることにより，平均比較回数 $n\lg n-1.4106n+O(\log n)$ が達成されることを示した．

1 はじめに

ソーティングは計算機における最も基本的な操作の一つである．多くのソーティングアルゴリズムは比較ソートであり，その計算効率は比較回数で計られる．計算時間の場合は正確な評価は困難で通常ビッグオー表現で定数係数や低次の項を無視する．しかし，比較回数ならより正確な評価をすることも可能であろう．実際，情報理論的下限が簡単に求まり，それは $\lceil \lg n! \rceil \approx n\lg n-1.4426n+O(\log n)$ で与えられる．問題は n の線形の項の係数で，これを上下限と一致させることが出来ないと信じられているようである．そこでその定数係数を如何に 1.4426 に近付けることが出来るかが自然な研究課題になる．しかし，その解析は簡単ではなく，現在までに得られている結果は非常に限られている．最悪の比較回数に関して分かっているのは，1959 年に提案された**マージ挿入法** [2] の $n\lg n-1.3289n+O(\log n)$ が現在知られている最良の結果である．最悪比較回数を求めるのはかなりの難題であると予想される．

本稿では解析がより容易になると考えられる平均の比較回数を議論する．平均なら得意な入力によって比較回数が異常に増加するまれな例をある程度無視する事ができるのがその利点である．実際，二分探索による挿入を繰り返してソートする**二分挿入法** [4] と最悪比較回数では最も優れていると考えられているマージ挿入法 [2] に関して平均の比較回数が得られており，それぞれ，$n\lg n-1.3817n+O(\log n)$, $n\lg n-1.3999n+O(\log n)$ である [1]．しかし，n の線形の項の係数という観点からみれば，下限とのギャップはかなり大きい．そのギャップを埋めたいという理論的興味は自然なものであろう．

本稿では，二分探索法の手法を拡張することにより，平均比較回数 $n\lg n-1.4034n+O(\log n)$ が達成されることを示した．我々のアイデアは，2つの要素を効率よく挿入する操作を提案する

ところにある．また，既存結果で最適であるマージ挿入法と組み合わせることにより，平均比較回数 $n \lg n - 1.4106n + O(\log n)$ が達成されることを示した．この結果により，下限とのギャップを約25パーセント縮めることに成功した．アルゴリズムとしては計算量が $O(n \log n)$ のものを考える．いくらでも計算時間を使っていいのなら，長さ n から最適な決定木を計算できてしまうからである．

関連研究： Edelkamp と Weiß [1] によって，Steinhaus [4] による二分探索法，及び Ford と Johnson [2] によるマージ挿入法の平均比較回数が与えられている．

二分探索法 [4] は二分探索によって挿入済みの列に要素を挿入するアルゴリズムである．長さ $k-1$ の整列済み列に1つの要素を挿入する時の平均比較回数は

$$\lceil \lg i \rceil + 1 - \frac{2^{\lceil \lg k \rceil}}{k}$$

であり，二分探索法の平均比較回数は

$$\begin{aligned} B(n) &= \sum_{k=1}^{n} \left(\lceil \lg i \rceil + 1 - \frac{2^{\lceil \lg k \rceil}}{k} \right) \\ &= n \lg n + \left(1 - \lg p_n - \frac{1 + \ln(4p_n)}{p_n} \right) \\ &< n \lg n - 1.386n \end{aligned}$$

である [1]．ここで，$p_n = \frac{n}{2^{\lceil \lg n \rceil}}$ を導入した．p_n は区間 $(0.5, 1]$ の値を取り，n と $2^{\lceil \lg n \rceil}$(n 以上かつ最小の2冪) の比を表すものである．

マージ挿入法 [2] は最悪比較回数及び平均比較回数において現在知られている最良のアルゴリズムである．特に $n \leq 15$ に対して最適な最悪比較回数でソーティングを行う．Edelkamp と Weiß [1] によって，マージ挿入法の平均比較回数は高々$n \lg n - 1.3999n + O(\lg n)$ であることが知られている．

2　2要素挿入の平均比較回数

我々のアルゴリズムは二分挿入法を拡張して，整列済み列に2つの要素を挿入する操作を行う．1つの要素に対する二分探索による挿入はそれ自体は最適である．しかし，整列済みの列長によっては比較回数に1の差が出てしまう事が避けられない．ソートでは n 回の挿入をおこなうので，この様な1回ごとの1回の比較回数の差が積み重なって大きな差になってしまう．それを防ぐには1回の挿入のアンバランスを減らす事と挿入の回数を減らすことが重要である．2要素挿入はこれらの要求に答えれくれる様に見える．勿論，単純な挿入が良い列長に対してはそれを使えば良い．

以下では，整列済みで長さ $i-2$ の列 $T = (t_1, \ldots, t_{i-2})$ に2つの要素を挿入することを考える．大まかな方針は以下のとおりである．

Step 1: 挿入したい2つの要素を比較する．小さいほうを A，大きいほうを B と呼ぶ．

Step 2: 要素 A の挿入位置を以下のように決定する．$x := \frac{2-\sqrt{2}}{2}, r := 1$ とする．また，関数 $\alpha(r)$ を $\alpha(r) := 1 - (1-x)^r$ と定義する．

Step 2-(1): $t_{\alpha(r) \cdot i}$ と比較する．もし，A のほうが小さければ (2) に移る．そうでなければ，$r := r+1$ とし，(1) に戻る．

Step 2-(2): 要素 A の挿入位置を $t_{\alpha(r-1) \cdot i}$ から $t_{\alpha(r) \cdot i}$ の間で偏り付き二分探索（後述）を行うことで決定する．

Step 3: 要素 B の挿入位置を A の挿入位置から最後までの間で二分探索を行うことによって決定する．

Step 1 では明らかに 1 回の比較を必要とする．次節より，Step 2-(2) で行う偏り付き二分探索の導入をし，Step 2 の平均比較回数を与える．その後，Step 3 の平均比較回数を解析し，本操作の全体の平均比較回数を与える．

2.1 Step 2 の解析

要素 A について $\mathbf{Pr}[A=k]=\frac{i-k}{\binom{i}{2}}$ であることから，$k<k'$ に対して，$\mathbf{Pr}[A=k]>\mathbf{Pr}[A=k']$ が成り立つ．また，Step 2 では要素 A の挿入位置が $t_{\alpha(r-1)*i}$ から $t_{\alpha(r)*i}$ の間であることが分かっている．二分探索の幅は

$$(\alpha(r)-\alpha(r-1))\cdot i=(1-x)^{r-1}x\cdot i$$

であり，この値を w_r とする．Step 2 で行う偏り付き二分探索とは，A の挿入位置としてとり得る w_r 通りの値のうち，小さいほう ℓ 通りで比較回数 $\lceil\lg w_r\rceil-1$，残りの $w_r-\ell$ 通りについて比較回数 $\lceil\lg w_r\rceil$ 回となる二分探索である．ここで $\ell=2^{\lceil\lg w_r\rceil}-w_r$ である．偏り付き二分探索について以下の補題が得られる．

補題 2.1. *Step 2-(1)* の比較回数が r である時，*Step 2-(2)* の平均比較は

$$\lceil\lg w_r\rceil+7-4\sqrt{2}-(10-6\sqrt{2})\cdot\frac{1}{p_r}+\frac{3-2\sqrt{2}}{p_r^2}$$

である．

さて，Step 2-(1) の比較回数が r である確率 $q(r)$ が $\frac{1}{2^r}$ に十分近いことから，$\lceil\lg w_r\rceil,\frac{1}{p_r},\frac{1}{p_r^2}$ の期待値を与えられる．さらに，以下の補題が得られる．

補題 2.2. *Step 2* の平均比較回数は，

$$\lceil\lg i\rceil+7-4\sqrt{2}-\frac{2^{\lceil\lg i\rceil}}{i}+\frac{1}{6}\cdot\frac{4^{\lceil\lg i\rceil}}{i^2}+\begin{cases}-\frac{1}{6}\cdot\frac{2^{\lceil\lg i\rceil}}{i}-\frac{1}{16}\cdot\frac{4^{\lceil\lg i\rceil}}{i^2}-\frac{2}{3} & \frac{i}{2^{\lceil\lg i\rceil}}\in(1/2,\frac{1+\sqrt{2}}{4}]\\ -\frac{\sqrt{2}}{3}\cdot\frac{2^{\lceil\lg i\rceil}}{i}-\frac{1}{3} & \frac{i}{2^{\lceil\lg i\rceil}}\in(\frac{1+\sqrt{2}}{4},\frac{2+\sqrt{2}}{4}]\\ -\frac{4}{3}\cdot\frac{2^{\lceil\lg i\rceil}}{i}+\frac{1}{4}\cdot\frac{4^{\lceil\lg i\rceil}}{i^2}+\frac{1}{3} & \frac{i}{2^{\lceil\lg i\rceil}}\in(\frac{2+\sqrt{2}}{4},1]\end{cases}$$

である．

2.2 Step 3 の解析

要素 B の挿入に関して以下の補題を与える．

補題 2.3. *Step 3* での平均比較回数は

$$\lceil\lg(i-1)\rceil+1-\frac{2}{p_i}+\frac{1}{3p_i^2}+O(1/i)$$

である．

Proof. 要素 A が $t_{\ell-1} < A < t_\ell$ を満たすとする．この時，要素 B の挿入位置は $i-\ell$ 通りであり，要素 B に関する二分探索の平均比較回数は $\lceil \log_2(i-\ell) \rceil + 1 - \frac{2^{\lceil \log_2(i-\ell) \rceil}}{i-\ell}$ となる．要素 A が $t_{\ell-1} < A < t_\ell$ となる確率は $\frac{i-\ell}{\binom{i}{2}}$ であることに注意すると，Step 3 の平均比較回数は

$$\sum_{\ell=1}^{n-1} \left\{ \lceil \log_2(i-\ell) \rceil + 1 - \frac{2^{\lceil \log_2(i-\ell) \rceil}}{i-\ell} \right\} \cdot \frac{i-\ell}{\binom{i}{2}} = 1 + \frac{\sum_1^{i-1} \{ t \lceil \lg t \rceil - 2^{\lceil \lg t \rceil} \}}{\binom{i}{2}}$$

となる．

$m = \lceil \lg(i-1) \rceil$ としてそれぞれの項の和をとると，

$$\sum_{t=1}^{i-1} t \lceil \lg t \rceil = m \binom{i}{2} - \frac{1}{6} 4^m - \frac{1}{2} 2^m + \frac{2}{3},$$

$$\sum_{t=1}^{n-1} 2^{\lceil \lg t \rceil} = i \cdot 2^m - \frac{1}{3} 4^m - 2^m - \frac{1}{6}$$

となる．よって

$$\frac{\sum_{t=1}^{i-1} \{ t \lceil \log_2 t \rceil - 2^{\lceil \log_2 t \rceil} \}}{\binom{i}{2}} = m - 2 \cdot \frac{2^m}{i} + \frac{1}{3} \cdot \frac{4^m}{i^2} + O(1/i)$$

であるから，補題が得られる． □

2.3 全体の平均比較回数

補題 2.2 と補題 2.3 より，整列済みの長さ $i-2$ の列に対する二要素挿入の平均比較回数は，

$$\lceil \lg i \rceil + \lceil \lg(i-1) \rceil + 9 - 4\sqrt{2} - \frac{3}{p_i} + \frac{1}{2p_i^2} + O(1/i) + \begin{cases} -\frac{1}{6p_i} - \frac{1}{16p_i^2} - \frac{2}{3} & p_i \in (1/2, \frac{1+\sqrt{2}}{4}] \\ -\frac{\sqrt{2}}{3p_i} - \frac{1}{3} & p_i \in (\frac{1+\sqrt{2}}{4}, \frac{2+\sqrt{2}}{4}] \\ -\frac{4}{3p_i} + \frac{1}{4p_i^2} + \frac{1}{3} & p_i \in (\frac{2+\sqrt{2}}{4}, 1] \end{cases}$$

となり，i 個目の要素を挿入するために

$$\lceil \lg i \rceil + \frac{9 - 4\sqrt{2}}{2} - \frac{3}{2p_i} + \frac{1}{4p_i^2} + O(1/i) + \begin{cases} -\frac{1}{12p_i} - \frac{1}{32p_i^2} - \frac{1}{3} & p_i \in (1/2, \frac{1+\sqrt{2}}{4}] \\ -\frac{\sqrt{2}}{6p_i} - \frac{1}{6} & p_i \in (\frac{1+\sqrt{2}}{4}, \frac{2+\sqrt{2}}{4}] \\ -\frac{2}{3p_i} + \frac{1}{8p_i^2} + \frac{1}{6} & p_i \in (\frac{2+\sqrt{2}}{4}, 1] \end{cases}$$

の平均比較回数が必要であるとみなせる．二分挿入法における二分探索の平均比較回数 $\lceil \lg i \rceil + 1 - \frac{1}{p_i}$ と比較すると，我々の手法は $0.5511 < p_i < 0.888$ を満たす i で効率よく挿入が行われる．

3 全体のアルゴリズム

2 節で整列済みの列長 i に対して，$0.5511 < p_i < 0.888$ を満たすならば，2 つの要素を挿入したほうがより少ない平均比較回数で挿入が行われることがわかった．我々のアルゴリズムは以下の操作を繰り返し行うことによりソーティングを行う．

[操作] 整列済みの列長 $i-2$ に対して，$0.5511 < p_i < 0.888$ を満たすならば，2 つの要素を挿入する方法を呼び出す．そうでなければ，二分探索によって 1 つの要素を挿入する．

この操作における i 個目の要素挿入に関する平均比較回数は以下のように表される．

$$\lceil \lg i \rceil + 1 - \frac{1}{p_i} + O(1/i) + \begin{cases} 0 & p_i \in (1/2, 0.5511] \\ \frac{19}{6} - 2\sqrt{2} - \frac{7}{12p_i} + \frac{7}{32p_i^2} & p_i \in (0.5511, \frac{1+\sqrt{2}}{4}] \\ \frac{10}{3} - 2\sqrt{2} - \frac{3+\sqrt{2}}{6p_i} + \frac{1}{4p_i^2} & p_i \in (\frac{1+\sqrt{2}}{4}, \frac{2+\sqrt{2}}{4}] \\ \frac{11}{3} - 2\sqrt{2} - \frac{7}{6} \cdot \frac{1}{p_i} + \frac{3}{8p_i^2} & p_i \in (\frac{2+\sqrt{2}}{4}, 0.888] \\ 0 & p_i \in (0.888, 1] \end{cases}$$

$B(n) = \sum_{i=1}^{n} (\lceil \lg i \rceil + 1 - \frac{1}{p_i})$ 及び $\sum_{i=1}^{n} O(1/i) = O(\log n)$ であることから，残りの項の和 $C(n)$ を考える．関数 $D(x)$ を残りの項に関して $x = p_i$ に置き換えた関数と定義する．

$$D(x) = \begin{cases} 0 & x \in (1/2, 0.5511] \\ \frac{19}{6} - 2\sqrt{2} - \frac{7}{12x} + \frac{7}{32x^2} & x \in (0.5511, \frac{1+\sqrt{2}}{4}] \\ \frac{10}{3} - 2\sqrt{2} - \frac{3+\sqrt{2}}{6x} + \frac{1}{4x^2} & x \in (\frac{1+\sqrt{2}}{4}, \frac{2+\sqrt{2}}{4}] \\ \frac{11}{3} - 2\sqrt{2} - \frac{7}{6} \cdot \frac{1}{x} + \frac{3}{8x^2} & x \in (\frac{2+\sqrt{2}}{4}, 0.888] \\ 0 & x \in (0.888, 1] \end{cases}$$

この時，

$$\begin{aligned} C(n) &\approx \sum_{d=1}^{\lceil \lg n \rceil - 1} 2^d \times \int_{1/2}^{1} D(x)dx + 2^{\lceil \lg n \rceil} \times \int_{1/2}^{p_n} D(x)dx \\ &\approx 2^{\lceil \lg n \rceil} \times \left\{ \int_{1/2}^{1} D(x)dx + \int_{1/2}^{p_n} D(x)dx \right\} \end{aligned}$$

となる．全体の平均比較回数は $B(n) + C(n) + O(\log n)$ であり，高々$n \lg n - 1.40118n$ である（図 1 参照）．

4 Step2 の改良

本節ではアルゴリズムの Step 2，つまり A の挿入に関して比較する要素を以下のように設定することで平均比較回数が改良できることを示す．もし $p_i \in (3/4, 1]$ ならば，関数

$$\alpha(r) = \begin{cases} 1 - \frac{1}{2^{k-1}} + \frac{1}{p_i 2^{k+1}} & (r = 2k-1) \\ 1 - \frac{1}{2^k} & (r = 2k) \end{cases}$$

と定義する．もし $p_i \in (1/2, 3/4]$ ならば，$\alpha(1) = \frac{1}{2} - \frac{1}{8p_i}$ とする．$t_{\alpha(1)i}$ の比較で A のほうが大きいとなった場合 $p_{(1-\alpha(1))i} = p_i + 1/4 \in (3/4, 1]$ を満たすため，$p_i \in (3/4, 1]$ の場合を適用することができる．

4.1 $p_i \in (3/4, 1]$ の場合

偏り付き二分探索の幅 $w_r := (\alpha(r) - \alpha(r-1))i$ は，

$$w_r = \begin{cases} \frac{i}{p_i \cdot 2^{k+1}} = \frac{2^{\lceil \lg i \rceil}}{2^{k+1}} & (r = 2k-1) \\ \frac{i}{2^k} \cdot \left(1 - \frac{1}{2p_i}\right) & (r = 2k) \end{cases}$$

となる．また，$\lg i = \lceil \lg i \rceil + \lg p_i$ かつ $p_i \in (0.75, 1]$ であることから，

$$\lceil \lg w_{2k-1} \rceil = \lceil \lg w_{2k} \rceil = \lceil \lg i \rceil - k - 1$$

である．さらに，

$$\mathbf{Pr}[r = 2k-1] = \frac{8p_i - 1}{p_i^2 4^{k+1}},\ \mathbf{Pr}[r = 2k] = \frac{(6p_i+1)(2p_i-1)}{4^{k+1}p_i^2}$$

を満たし，

$$\sum_{k \geq 1} \mathbf{Pr}[r = 2k] = \frac{(6p_i-1)(2p_i-1)}{12p_i^2}$$

である．

$r = 2k-1$ の時，二分探索にかかる比較回数は幅 w_{2k-1} が 2 冪であることから $\lceil \lg i \rceil - k - 1$ となる．$r = 2k$ の時は 3 節と同様にして，

$$\lceil \lg i \rceil - k - 2 + \frac{32p_i^2 - 36p_i + 9}{(2p_i-1)(6p_i-1)}$$

が得られる．以上から Step 2 にかかる平均比較回数は

$$\begin{aligned}
& \sum_{k \geq 1} \mathbf{Pr}[r = 2k-1] \times \{2k - 1 + \lceil \lg i \rceil - k - 1\} \\
& + \sum_{k \geq 1} \mathbf{Pr}[r = 2k] \times \left\{2k + \lceil \lg i \rceil - k - 2 + \frac{32p_i^2 - 36p_i + 9}{(2p_i-1)(6p_i-1)}\right\} \\
= \ & \lceil \lg i \rceil - 2 + \sum_{k \geq 1} \mathbf{Pr}[r = 2k-1] \times k \\
& + \sum_{k \geq 1} \mathbf{Pr}[r = 2k] \times \left\{k + \frac{32p_i^2 - 36p_i + 9}{(2p_i-1)(6p_i-1)}\right\} \\
= \ & \lceil \lg i \rceil - 2 + \sum_{k \geq 1} \frac{3k}{4^k} + \frac{(2p_i-1)(6p_i-1)}{12p_i^2} \times \frac{32p_i^2 - 36p_i + 9}{(2p_i-1)(6p_i-1)} \\
= \ & \lceil \lg i \rceil + 2 - \frac{3}{p_i} + \frac{3}{4p_i^2}
\end{aligned}$$

4.2 $p_i \in (1/2, 3/4]$ の場合

$\alpha(1) = \frac{1}{2} - \frac{1}{8p_i}$ より $\lceil \lg(\alpha(1)i) \rceil = \lceil \lg i \rceil - 2$ である．

$r = 1$ に対する偏り付き二分探索の比較回数は，

$$\begin{aligned}
& \lceil \lg \alpha(1)i \rceil - \left(\frac{2^{\lceil \lg \alpha(1)i \rceil} - \alpha(1)i}{\alpha(1)i}\right) \cdot \left(1 + \frac{2\alpha(1)i - 2^{\lceil \lg \alpha(1)i \rceil}}{\alpha(1)i} \cdot \frac{\alpha(1)}{2 - \alpha(1)}\right) \\
= \ & \lceil \lg i \rceil + \frac{-16p_i^2 - 56p_i + 11}{(4p_i-1)(12p_i+1)}
\end{aligned}$$

である．ここで $T = \frac{-16p_i^2 - 56p_i + 11}{(4p_i-1)(12p_i+1)}$ とする．また，

$$\mathbf{Pr}[r = 1] = \frac{(4p_i-1)(12p_i+1)}{64p_i^2},\ \mathbf{Pr}[r > 1] = \frac{(4p_i+1)^2}{64p_i^2} = \frac{(p_i + 1/4)^2}{4p_i^2}$$

である．また，$r>1$ である場合の平均比較回数は，

$$\lceil \lg i \rceil + 1 - \frac{3}{p_i+1/4} + \frac{3}{4(p_i+1/4)^2}$$

で与えられる．よって，Step 2 全体の平均比較回数は，

$$\begin{aligned}
& \lceil \lg i \rceil + 1 + \mathbf{Pr}[r=1] \times T + (1-\mathbf{Pr}[r=1]) \times \left(1 - \frac{3}{p_i+1/4} + \frac{3}{4(p_i+1/4)^2}\right) \\
= & \lceil \lg i \rceil + 1 + \frac{-16p_i^2 - 56p_i + 11}{64p_i^2} + \frac{(4p_i+1)^2}{64p_i^2} - \frac{3p_i+3}{16p_i^2} + \frac{3}{16p_i^2} \\
= & \lceil \lg i \rceil + 1 - \frac{3}{2p_i} + \frac{3}{16p_i^2}
\end{aligned}$$

となる．

4.3 全体の平均比較回数

改良した二要素挿入における i 個目の要素を挿入するための平均比較回数は，

$$\lceil \lg i \rceil + 1 - \frac{1}{p_i} + \begin{cases} \frac{1}{2} - \frac{3}{4p_i} + \frac{25}{96p_i^2} & p_i \in (1/2, 3/4] \\ 1 - \frac{3}{2p_i} + \frac{13}{24p_i^2} & p_i \in (3/4, 1] \end{cases}$$

となる．後ろの項を考えると，$p_i \in \left(\frac{3}{4} - \frac{\sqrt{6}}{12}, \frac{3}{4} + \frac{\sqrt{3}}{12}\right)$ で通常の二分探索よりも効率的であることが分かる．さて，関数 $D'(x)$ を以下のように定義する．

$$D'(x) = \begin{cases} 0 & (1/2, \frac{3}{4} - \frac{\sqrt{6}}{12}] \\ \frac{1}{2} - \frac{3}{4x_i} + \frac{25}{96x^2} & x \in (1/2, 3/4] \\ 1 - \frac{3}{2x_i} + \frac{13}{24x_i^2} & x \in (3/4, \frac{3}{4} + \frac{\sqrt{3}}{12}] \\ 0 & x \in (\frac{3}{4} + \frac{\sqrt{3}}{12}, 1] \end{cases}$$

3 節と同様に

$$C'(n) = 2^{\lceil \lg n \rceil} \times \left\{ \int_{1/2}^{1} D'(x)dx + \int_{1/2}^{1} D'(x)dx \right\}$$

とすると，アルゴリズム全体の平均比較回数は $B(n) + C'(n) + O(\log n)$ で表され，全体の平均比較回数は高々 $n \lg n - 1.4034n$ である（図 1 参照）．

5 マージ挿入法との併用

マージ挿入法は最悪比較回数においても現在知られている最良のアルゴリズムである．ある k に対して $n = \lceil \frac{2^k}{3} \rceil$ である時，マージ挿入法の最悪比較回数は高々 $n \lg n - (3 - \lg 3)n + O(\lg n) \approx n \lg n - 1.415n + O(\lg n)$ である．

本節では，以下のアルゴリズムを考える．

[併用アルゴリズム] 入力列の長さ n に対して，$\frac{2^k}{3} < n$ を満たす最大の k を選び，$n' := \frac{2^k}{3}$ とする．n' 個の要素についてマージ挿入法を適用し整列済みの列を得る．長さ n になるまで前節で改良した挿入を行う．

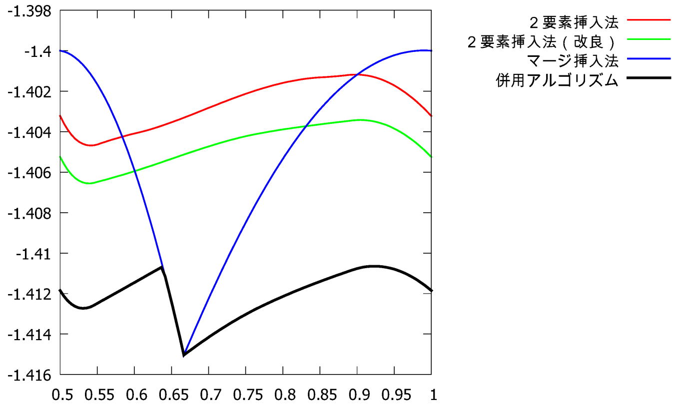

図 1: 平均比較回数 $n\lg n + cn$ に対する c 値

$p_n \geq 2/3$ の場合，$n' = \frac{2n}{3p_n}$ である．この時，マージ挿入法でかかる平均比較回数は

$$n'\lg n' - (3 - \lg 3)n' = n'\lceil \lg n \rceil - \frac{4}{3p}n$$

である．また，整列済み長さ n' 列への挿入に関する平均比較回数は

$$\sum_{n'+1}^{n} \lceil \lg i \rceil + 2^{\lceil \lg n \rceil} \int_{2/3}^{p_n} D'(x)dx = (n - n')\lceil \lg n \rceil + 2^{\lceil \lg n \rceil} \int_{2/3}^{p_n} D'(x)dx$$

となる．よって，全体の平均比較回数は

$$\begin{aligned} & n'\lceil \lg n \rceil - \frac{4}{3p}n + (n - n')\lceil \lg n \rceil + 2^{\lceil \lg n \rceil} \int_{2/3}^{p} D(x)dx \\ = \; & n\lceil \lg n \rceil - \frac{4}{3p}n + 2^{\lceil \lg n \rceil} \int_{2/3}^{p} D(x)dx \end{aligned}$$

で求まる．$p_n \leq 1/3$ の場合も同様にして平均比較回数を求めることができる．各 n において Edelkamp と Weiß [1] が示したマージ挿入法の平均比較回数のよいほうを選択すると，平均比較回数は高々 $n\lg n - 1.41064n$ となる（図 1 参照）．

参考文献

[1] S. Edelkamp and A. Weiß. QuickXsort: Efficient Sorting with $n\log n - 1.399n + o(n)$ Comparisons on Average. CSR 2014: pp. 139–152.

[2] J. Ford, Lester R. and S. M. Johnson. A tournament problem. The American Mathematical Monthly, 66(5):pp. 387–389, 1959.

[3] D. E. Knuth. Sorting and Searching, volume 3 of The Art of Computer Programming. Addison Wesley Longman, 2nd edition, 1998.

[4] H. Steinhaus, Mathematical Snapshots, New Nork, 1950, pp. 37–40.